Phylogenetic Analysis of RSV Strains

By: Carolina Parra and Shyan Polman

Introduction

Human respiratory syncytial virus (RSV), a common respiratory virus, has presented a growing concern in clinical settings, particularly with infant and child infections. 90% of infants are infected with RSV within the first two years of their lives, resulting in reinfections in older children[1]. RSV typically infects the upper respiratory system and causes cold-like symptoms, such as runny nose, sore throat, and low-grade fever, however, severe symptoms can occur in infants, such as difficulty breathing, severe cough, and poor feeding[2]. There currently are two known RSV strains: RSV-A and RSV-B. Between these two strains, RSV-A is known to be more prevalent, contagious, and fatal[3]. The phylogenetic analysis between these two strains will further help understand why the RSV-A strain is more dangerous to our society. With this analysis, we can dive deeper to observe any differences, such as genetic or structural.

Questions

How do the phylogenetic relationships between RSV-A and RSV-B correlate with their structural disparities?

How do these differences contribute to their infection and ability to evade immune responses?

Methods

The data sequence files for each stain were collected from the public NCBI Virus database. A total of 200 random sequence files per strain were collected. Following the collection, sequence files were aligned using multiple sequence alignment MAFFT on

JetStream2, and the resulting alignments were then visually inspected for quality. The maximum likelihood phylogenetic trees were constructed using Randomized Axelerated Maximum Likelihood (RAxML) algorithm. The aligned sequences were used as input for RAxML, with the GTR model of nucleotide substitution and the gamma distribution rate of heterogeneity, and the best scoring tree was selected for visualization. To view the trees, the Newick annotations were exported to the Interactive Tree of Life (ITOL) visualizer. The resulting trees were visually inspected to investigate the amount of genetic change between strains based on branch length. Protein sequence files corresponding to the primary sequences of interest from the phylogenetic tree were retrieved from the aligned nucleotide sequences. Using these sequence files, protein structures were predicted using the cloud-based ColabFold[6] platform that uses AlphaFold2 and Alphafold2-multimer. The resulting predicted structures were then compared to see any significant differences between the strains.

Results

Figure 1: Maximum Likelihood Phylogeny Tree of RSV-A was conducted using RAxML with 200 sequence files

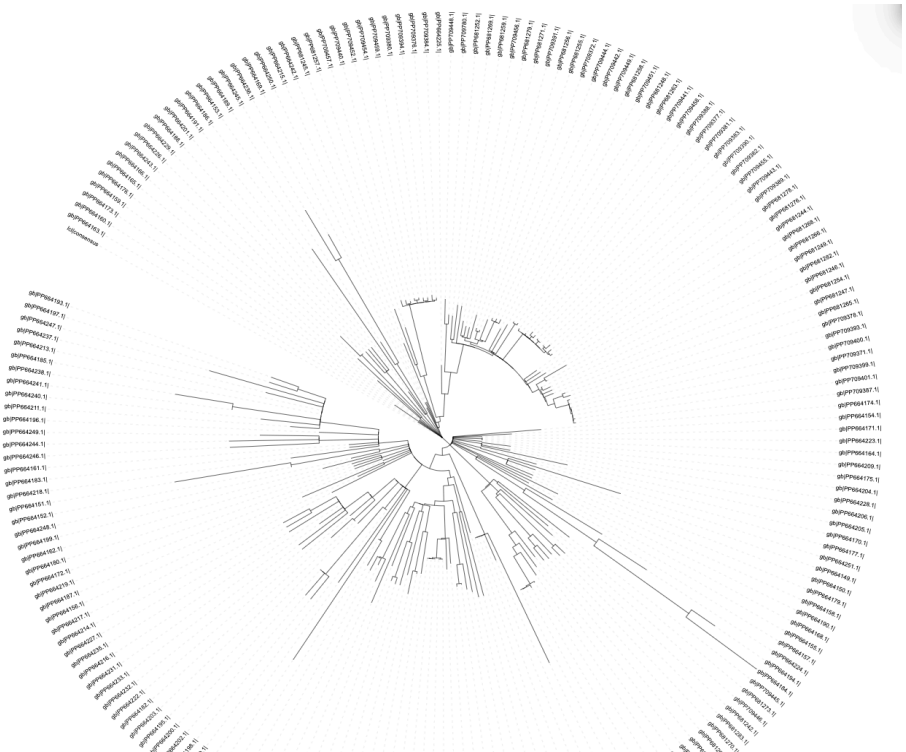

Figure 2: Maximum Likelihood Phylogeny Tree of RSV-B was conducted using RAxML with 200 sequence files

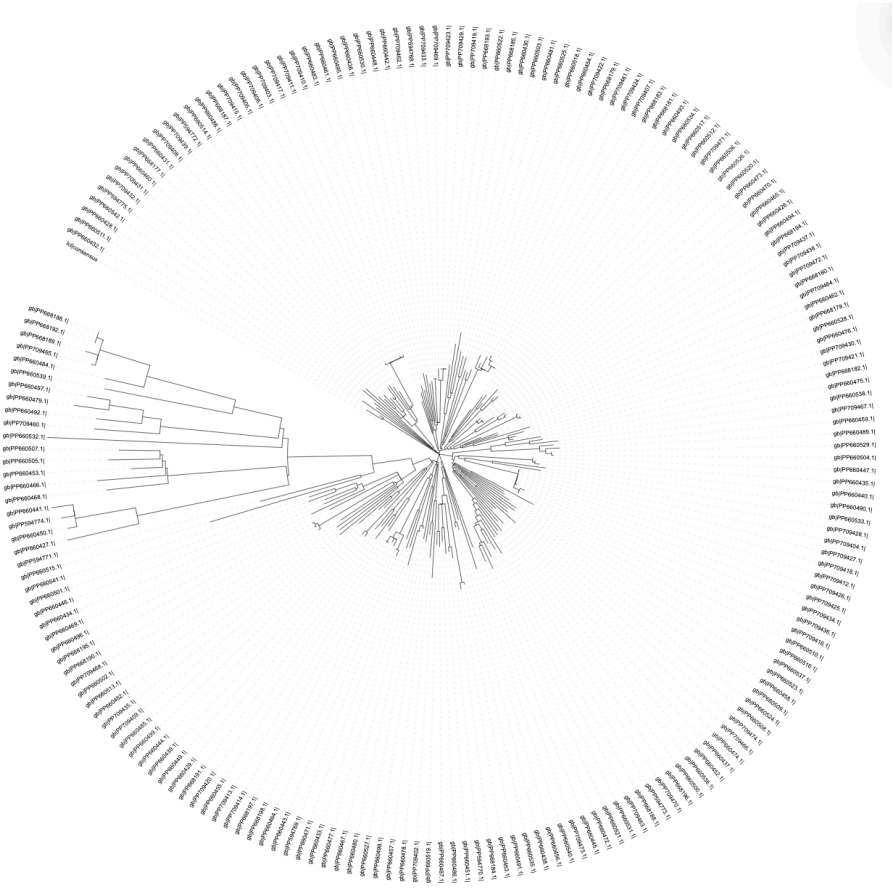

Figure 3: Maximum Likelihood Phylogeny Tree of RSV-A and RSV-B was conducted using RAxML with 400 sequence files

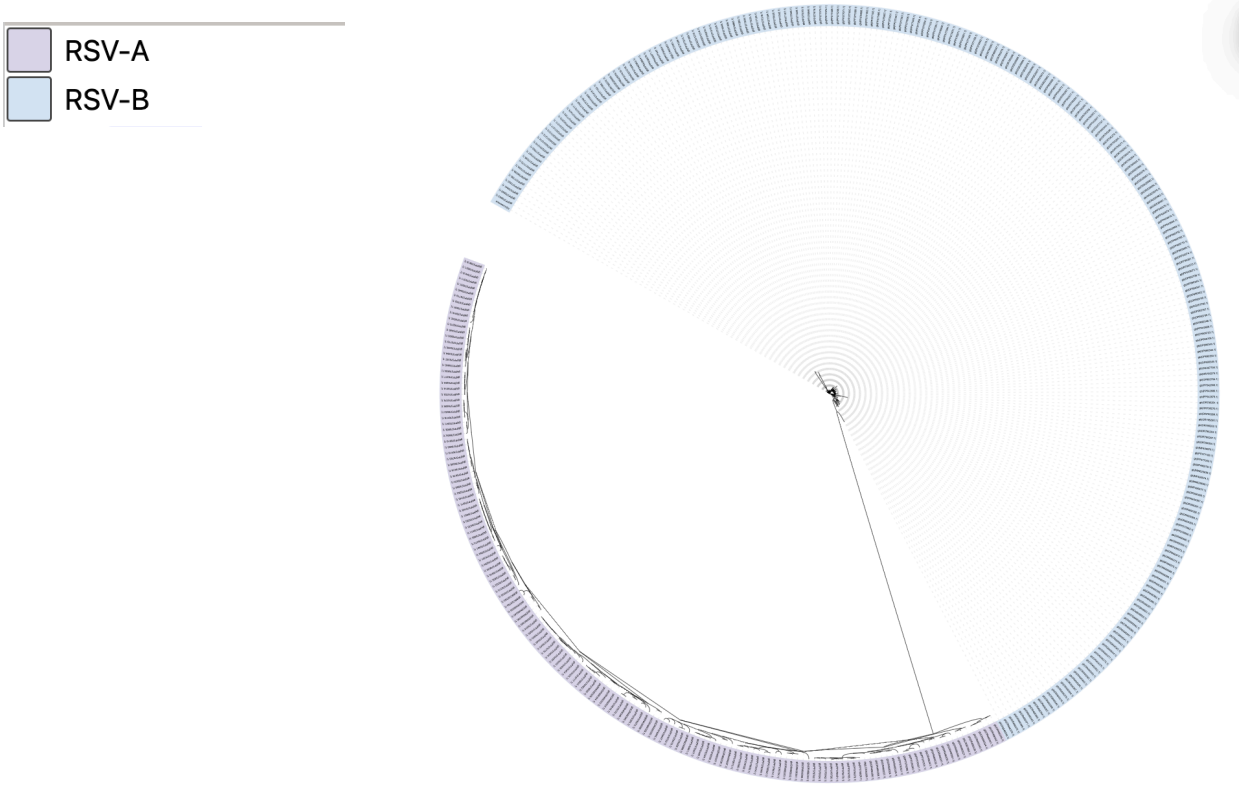

Figure 4: Predicted glycoprotein structure of RSV-A using ColabFold

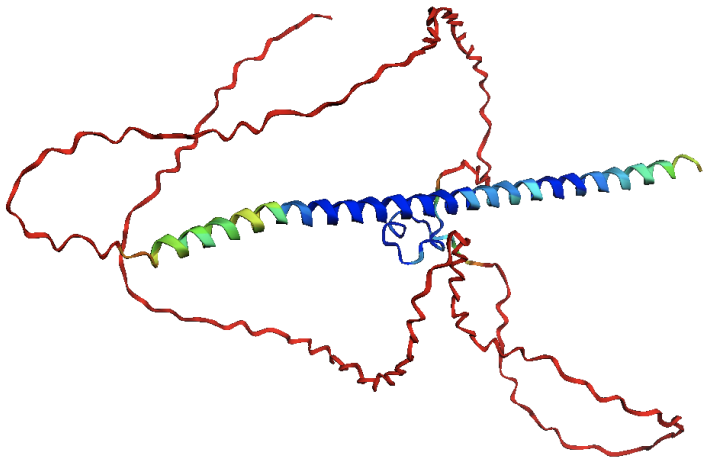

Figure 5: Predicted glycoprotein structure of RSV-B using ColabFold

Discussion

The longer branch lengths in the RSV-A phylogenetic tree (Figure 1) compared to the phylogenetic tree of RSV-B (Figure 2) indicate higher genetic diversity and more distanced evolutionary clades. Higher genetic diversity can lead to an increase of genetic mutation and variation within the RSV-A strain than RSV-B[5]. Further examination of the combined phylogenetic tree displayed a sharp evolutionary difference between strain A and B (Figure 3). The phylogenetic differences between the two strains prompted further analysis of the structural differences, specifically the glycoprotein used to initiate infection. The RSV-A glycoprotein (Figure 4) resulted in containing a longer alpha helix compared to that of RSV-B (Figure 5). This may implicate certain contributions to the stability and efficiency of the glycoprotein structure and function in RSV-A. Further investigative analysis to prove this can be done through testing the efficiency of the two glycoproteins in ELISA binding assays and analyzing its interactions

with host cell receptors. Understanding their efficiency and mechanisms of infection can eventually lead to improvement of antiviral drugs or vaccines for RSV. Overall, the phylogenetic relationships between RSV-A and RSV-B can be correlated to their structural disparities, thus may impact their ability to invade susceptible hosts.

References

1. https://www.ncbi.nlm.nih.gov/books/NBK459215/
2. https://www.mayoclinic.org/diseases-conditions/respiratory-syncytial-virus/symptoms-causes/syc-20353098
3. https://www.nature.com/articles/s41598-023-40760-y#:~:text=Interestingly%2C%20RSVA%20is%20more%20common,it%20causes%20an%20infection8.
4. https://pubmed.ncbi.nlm.nih.gov/27741144/
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5147804/
6. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb